DRAFT

Comparative Cognitive Task Analysis

Susan S. Kirschenbaum[1]

Naval Undersea Warfare Center Division Newport

J. Gregory Trafton

Naval Research Laboratory

Elizabeth A. Kramer

University of Connecticut

Approval Pending:

Approved for public release: Distribution is unlimited

[1] Address correspondence to Dr. Susan S. Kirschenbaum, Naval Undersea Warfare Center Division Newport, 1176 Howell St., Code 2211, Building 1171/1, Newport, RI 02841-1708, email: kirschenbaumss@npt.nuwc.navy.mil.

Abstract

Cognitive Task Analysis (CTA) is a recognized technique for analyzing how people operate within a specific task domain and with a specific toolset. In this paper we introduce an improved variant, Comparative CTA (C2TA) that allows the separation of effects due to basic human cognition from those due to the toolset. An example from weather forecasting is provided, including sample results and recommendations that would not have emerged from traditional CTA.

Intro

Weather forecasting is a complex process. The supporting information is multi-dimensional, distributed, and often uncertain. It includes both “raw” observations (e.g., current temperature, winds, pressure, clouds, precipitation, radar returns, satellite pictures, etc.) and analytic weather models, at varying scales predicting future conditions. How does the designer incorporate User-Centered Design and Human Centered Computing into this complex and specialized domain? How does the designer gain enough knowledge of the users’ tasks and processes to provide useful assistance? And how does the designer disentangle the effects of task, training, teamwork arrangements, and basic human cognition from those of the design of the tools?

The traditional way human factors engineers approach this problem is to perform a task analyses to determine how people operate in a specific domain on a specific task. Cognitive Task Analysis (CTA) is a set of methods that takes into account the perception (i.e., vision), cognition (i.e., decision making), and motor actions (i.e., mouse movements) to accomplish a task.

In this paper, we build on traditional CTA methods by suggesting comparative cognitive task analysis can help solve the above problems. Comparative CTA (C2TA) is based on replication studies conducted in different environments. Because it derives data from more than one environment, C2TA provides insight into interface design that single site studies and single CTA methods cannot.

There are many versions of task analysis ranging from time and motion study [1] to GOMS (Goals, Operators, Methods, Selection rules) analysis [2] to Ecological Interface Design (EID) [3]. Each is best suited to a subset of design problems. For example, GOMS analysis is a keystroke-level process for describing human-computer interactions (e.g., mouse and keyboard interactions). EID focuses on how the operator interacts with indicators of physical functioning such as in a power plant or manufacturing control room. CTA is especially useful in situations where the task is heavily dependant on human interpretation and integration of dynamic and highly uncertain data [4]. Weather forecasters typically deal with large amounts of data over time and space. Additionally, the information they examine is uncertain on several dimensions (i.e., the predictive weather models that are run may be based on a small number of data points in some areas (like in the middle of the ocean), which necessitates interpolating from the current data, which may cause the final output to be more uncertain). The dependence on a meteorologist interpreting the weather models, the dynamic nature of weather, and the uncertainty in the weather models makes weather forecasting an excellent candidate for CTA.

However, most of the data analyzed by CTA methods come from a single source (i.e., most CTA studies have been performed on a single system and/or a small group of people). While the single approach is adequate in many situations, it may not be as generalizable as it could be. That is, any problems can typically be traced to the interaction between the person and the system. You may discover, for example, that a specific pointing device is not very effective on a particular system, but you do not know if that is a limitation of the pointing device or the way people think; you only know that the combination of people and pointing device on the task you are examining is not very effective. By examining radically different tools (i.e., different types of pointing devices on similar tasks, you can start to dissociate the effects of cognition and tool. For example, the pen, the typewriter, and the computer keyboard are all tools that can be used for writing a document. The writing process consists of planning, composing, editing, and production (writing / typing). The quantity and sequence of these processes is differentially supported by the three tools. For example, the computer supports longer compositions, however, the writer plans longer before editing with a pen [5]. This may be because editing with a pen includes crossing out, rewriting, cutting pages apart and taping them back together, arrows for inserts, etc. and then repeating the production process (re-writing) on clean paper. Editing on a typewriter uses similar cross out, cut, glue, and retype processes. With both of these tools, the re-write (production) process is effortful. However, writers using a computer edit more as they write and new versions do not require re-doing the physical production [6].

The data for the two analyses reported here were collected during two studies in two very different locations, a United States Navy (USN) Meteorology and Oceanography (METOC) center in California and a Royal Australian Navy (RAN) METOC facility. These studies employed the methods of both naturalistic field study and quasi-naturalistic observation in a laboratory field study. The studies were part of a project to provide improved tools for Navy weather forecasting. Only by understanding current practices and forecasting tools could improvements be suggested that would make the process more efficient. The two studies allowed us to map the information usage of decision maker to information visualization tools and to compare the mappings of USN and RAN forecasters in order to distinguish between effects that are dictated by the tools and training of these specialists and those due to basic human cognition.

In the remainder of this paper we will first briefly describe the data collection at the two sites. Then we will review the results of the C2TA and show how suggestions for the design or redesign of tools flow from the C2TA results. More detailed results from both studies can be found in [7, 8].

Two Studies

USN (2000)

The first of the two studies took place in San Diego, CA at a Naval meteorological and oceanographic location. This laboratory field study provided naval forecasters in a simulated METOC center with computer access to their usual forecasting tools. Most of the forecasting information came from numerous meteorological web sites including military, non-military government, and university sites.

Three pairs, consisting of a forecaster and technician took part in the study. Each pair developed a forecast and prepared a forecast briefing for an air strike to take place 12 hours in the future on Whidbey Island, WA. All actions were video taped and the participants were requested to “talk aloud” so as to produce a verbal protocol.

RAN (2001)

The second study was a naturalistic observation of RAN forecasters working at a Weather and Oceanography Centre at an airbase in eastern Australia. They too were forecasting for 12 hour (and 24 to 72 hour) air operations, prepared forecasts and forecast briefings, and used computer-based tools. As with the USN forecasters, most of the forecasting information came from numerous meteorological web sites. Also, like the USN forecasters, they were video taped and instructed to “talk aloud” to produce verbal protocols.

By retaining the task (forecasting) and moving to another group of practitioners with different tools, training, and teamwork practices, we disentangle the effects due to human cognition versus effects due to the tools used. There-by replicating basic results, extending findings, and drawing conclusions about how to better support the common forecasting tasks for both groups.

Results

Information Usage

Comparative CTA can tell two kinds of stories. *Similarities* in classes of information usage that are independent of the tools, training, and teamwork patterns imply basic processes of human cognition. In contrast, we can impute *differences* in information usage patterns as being due to the impact of differences in tools, training and teamwork. To find either, we must code the verbal protocols to capture the way the forecasters use information. To analyze these data we selected usage encodings that capture what the forecaster did with the information. In other reports we have examined the format of the information (text, graph, animation, etc.) or the form of the information (qualitative or quantitative) [7, 8].

The major encoding categories are *extracting* information, *comparing* information from two or more visualizations, *deriving* information by combining what was available in the visualization with the forecaster's knowledge, and *recording* information. Extracting information occurs when a forecaster examines a visualization and extracts some sort of local or global features that are explicitly represented in the visualization. Comparing information occurs when a forecaster compares two or more different visualizations. Deriving information occurs when a forecaster goes beyond the information on the visualization and makes inferences or combines the displayed information with their background knowledge. *Deriving* implies that some of the information comes from the forecasters' general domain knowledge rather than the information about the specific conditions for the current forecast. Recording information occurs when the forecaster writes down or copies information for their weather prediction. Table 1 defines each and gives an example. Note that, in terms of expertise required and cognitive work, there is a clear ordering from simplest to most demanding: *Extract* < *compare* < *derive*.

Table 1. Coding scheme with examples of each. The examples come from USN transcripts.

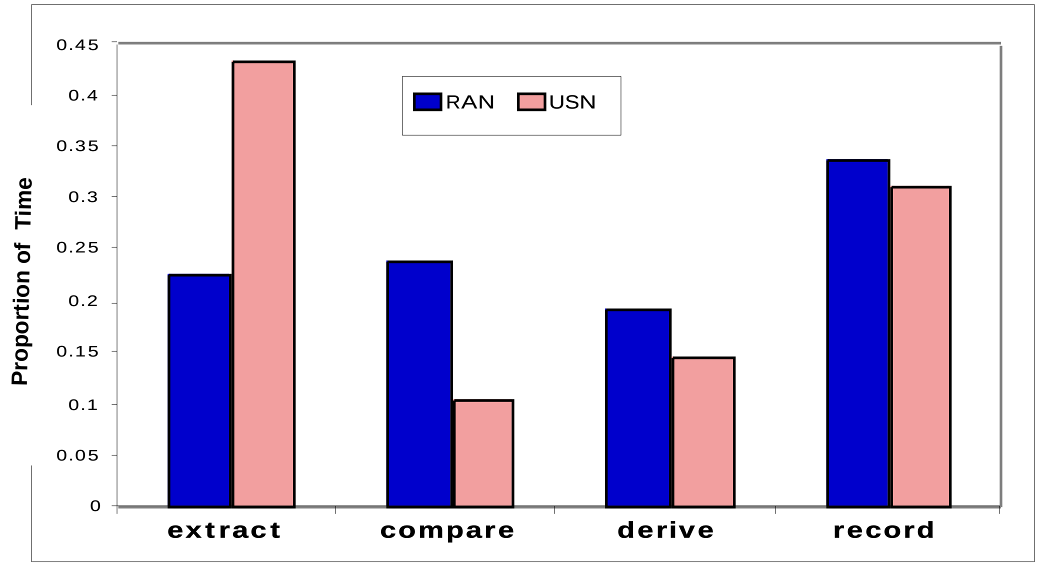
Usage	Definition	Example
Extract	To read information from any visible source.	looks like PVA over the area
Compare	To use two or more sources and comparing them on any data	radar shows precipitation, but I can't really see anything on the satellite picture.

	comparing them on any data	really see anything on the satellite picture.
Derive	To combine visible information with knowledge so as to come to a conclusion that is different from that which is in the visible source.	I think that's probably a little fast due to the fact that I don't think the models taking into account the topography of the area.
Record	Recording information for reporting to users. It need not be the final form.	This is a good picture right here, I'll take this….Just crop this picture a little bit.

The encodings were analyzed and the results compared between USN and RAN studies. They indicate a strong similarity between USN and RAN information usage. The basic processes are the same. There were no methods that were used by one group but not by the other. However, the order, tools used, and relative frequency with which these methods were used show significant differences in some areas. These areas are indications that the tool differently support the tasks. They are of interest for C2AT and for the information they provide about opportunities to improve the toolset.

Figure 1 indicates differences in the details of how USN and RAN forecasters accomplish their task, using the resources at hand and within their own specific environments (weather, training, and manning). We will concentrate on differences during the central tasks of developing and verifying the forecast. (There are no differences in the relative frequency of *record* actions even though specific tools and the pattern of tasks did differ.)

Figure 1. Proportion of time spent performing cognitive processes for all forecasters



Two observations stand out. The RAN forecasters appear to spend the same proportion of time in *extracting*, *comparing,* and *deriving* information while the USN forecasters spend far more of their time extracting information.

Compared to the RAN forecasters, the USN forecasters appear to spend a larger proportion of their time extracting information. In contrast, RAN forecasters spend virtually as much time *comparing* as *extracting* data. Thus, compared to the USN forecasters, the RAN forecasters spent a significantly larger proportion of their time engaged in comparing information.

C2TA reveals the differences between the two groups. However, the analyst must find the reasons for these differences. Candidate causes include task, tool, and training differences. In this case, the task is the same, predicting weather for naval aviation operations in the 12+ hour time frame. While training differs between the groups, tool differences appear to be the more likely cause. For example, the RAN forecasters have better support for comparisons because they either use adjacent monitors or adjacent windows on the same monitor. Thus, they can see a model and satellite or radar picture simultaneously or can examine two models side by side on the same monitor (Figure 2). In contrast, the USN forecasters must extract information from one data source, store it in memory or on paper, and then make comparisons from memory. With the RAN dual view (either on the same or adjacent monitors) the forecaster can make direct *comparisons*. The *extract* is an integral part of the process while the storage burden is greatly reduced. In Figure 2 the forecaster is comparing two models displayed side by side on the same monitor. Other comparisons observed are comparisons of predictions for the same model across time and comparisons of the model prediction for current time and current observations (e.g., a satellite picture on an adjacent monitor).

Figure 2. Forecaster comparing two models.

Sequences

Another C2TA observation from Figure 1 is that both groups spend a considerable portion of their time recording information for use in their forecasts. Further insight into this process can be achieved by examining the sequence of processes. Table 2 shows the probability of going from one process to another for both a sample USN and RAN forecaster. For example, given that the RAN forecaster is currently extracting data, his probability of his next action being *comparing, deriving,* or *recording* are p = 0.11, 0.44, and 0.44, respectively. In Table 2, these are represented by text weight with darker text representing higher probability of transition. This transition table emphasize the importance of the two poles, *extract* and *record.* These are the most common transition points for both the USN and RAN forecaster. Of the 3-node transitions, the most common cycles for both was either *extract* -> *record* -> *extract* or *record*-> *extract* -> *record*. For RAN, the *extract* -> *derive* -> *record* cycle was also common. Transitions between *compare* and *derive*, are noticeably fewer than those involving the poles.

As with the frequency data, sequence data provides insight into how tools do (or do not) support the cognitive tasks that make up weather forecasting. Design implications from the sequence data suggest the most effective places to automate. For example, as *extract* -> *record* sequences are common, a semi-automated tool might allow the forecaster who is *extracting* information to *record* the selected data at the press of a button and without changing screens. This would speed the recording process, eliminate accidental recording errors (typos, memory errors, etc.) and reduce the need to cycle between two tools.

Table 2. Probability of going from \ to process.

From \ To	Extract	Compare	Derive	Record
RAN				
Extract	0.00	0.11	**0.44**	**0.44**
Compare	**0.40**	0.00	0.20	**0.40**

Derive	0.31	0.08	0.00	**0.62**
Record	**0.48**	0.24	0.29	0.00
USN				
Extract	0.08	0.25	**0.42**	0.25
Compare	0.33	0.00	0.33	0.33
Derive	**0.50**	0.13	0.00	0.38
Record	**0.60**	0.00	0.20	0.20

Implications

With only traditional CTA, we could have observed the processes of *extraction, comparison*, *deriving*, and *recording* during the development of a weather forecast and forecasters cycle between developing their forecast (*extract*, *compare, derive*) and recording data. We would not have known that these processes and cycles are common to other forecasting environments. Furthermore, we would not have learned the important role that the supporting tools play in the *comparison* process.

The analyst and designer must work together to exploit these observations to guide the development of better tools. C2TA is only the first step but one that can inform and guide design toward making improvements where they are most needed.. However, the analyst and designer could develop tools to further facilitate the *comparison* process. . For example, as weather models are mathematical, they could be compared computationally with the results displayed in a single visualization. Agreements and disagreements could be highlighted. Models could be superimposed over satellite pictures for current model comparisons.

These are just examples of the kinds of conclusions that can be derived from C2TA. With a single data set, the designer can not know if the observed behavior is due to some demand characteristic of the toolset or to some facet of human cognition. With the addition of a second data set, the designer can separate the two and is thus free to develop better ways to support common cognitive processes with new tools.

Acknowledgements

Susan Kirschenbaum’s research was supported in part by the Office of Naval Research (ONR) as Project A424102 and by the Naval Undersea Warfare Center’s (NUWC) In-House Laboratory Independent Research (ILIR) Program, as Project E102077. The ILIR program is funded by ONR. Greg Trafton’s research was supported by Grants N00014-00-WX-20844 and N00014-00-WX-40002 from the Office of Naval Research.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U. S. Navy.

References

[1] F. B. Gilbreth and L. M. Gilbreth, *Applied motion study*. New York: The MacMillan Company, 1917.

[2] S. K. Card, T. P. Moran, and A. Newell, *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

[3] K. J. Vicente and J. Rasmussen, "Ecological Interface Design: Theoretical Foundations," *IEEE transactions on systems, man, and cybernetics*, vol. 22, pp. 586-606, 1992.

[4] J. M. Schraagen, S. F. Chipman, and V. L. Shalin, [Eds.], *Cognitive Task Analysis*. Mahwah, NJ: Erlbaum, 2000.

[5] C. Haas, "Does the Medium Make a Difference? Two Studies of Writing with Pen and Paper and with Computers," *Human-Computer Interaction*, vol. 4, pp. 149-169, 1989.

[6] R. T. Kellogg and S. Mueller, "Performance amplification and process restructuring in computer-based writing.," *International journal of man-machine studies*, vol. 39, pp. 33-49, 1993.

[7] J. G. Trafton, S. S. Kirschenbaum, T. L. Tsui, R. T. Miyamoto, J. A. Ballas, and P. D. Raymond, "Turning Pictures into Numbers: Use of Complex Visualizations," *International Journal of Human Computer Studies, Special issue on "Empirical Evaluations of Information Visualizations."*, vol. 53, pp. 827-850, 2000.

[8] S. S. Kirschenbaum, "Royal Australian Navy Meteorology and Oceanography Operations Research Report," Naval Undersea Warfare Center Division, Newport, RI, Technical Report 11,346, 25 January 2002.